Article:
Kennedy, H., Steedman, R. and Jones, R. (2021) Approaching public perceptions of datafication through the lens of inequality: a case study in public service media. Information, Communication and Society, 24 (12). pp. 1745-1761. ISSN 1369-118X

https://doi.org/10.1080/1369118X.2020.1736122

Routledge
Taylor & Francis Group

Information, Communication & Society

ISSN: 1369-118X (Print) 1468-4462 (Online) Journal homepage: https://www.tandfonline.com/loi/rics20

Approaching public perceptions of datafication through the lens of inequality: a case study in public service media

Helen Kennedy, Robin Steedman & Rhianne Jones

To cite this article: Helen Kennedy, Robin Steedman & Rhianne Jones (2020): Approaching public perceptions of datafication through the lens of inequality: a case study in public service media, Information, Communication & Society, DOI: 10.1080/1369118X.2020.1736122

To link to this article: https://doi.org/10.1080/1369118X.2020.1736122

Published online: 04 Mar 2020.

INFORMATION, COMMUNICATION & SOCIETY
https://doi.org/10.1080/1369118X.2020.1736122

Routledge
Taylor & Francis Group

Approaching public perceptions of datafication through the lens of inequality: a case study in public service media

Helen Kennedy[a], Robin Steedman[a] and Rhianne Jones[b]

[a]Department of Sociological Studies, University of Sheffield, Sheffield, UK; [b]BBC R&D, Salford, UK

ABSTRACT

In the emerging field of critical data studies, there is increasing acknowledgement that the negative effects of datafication are not experienced equally by all. Research on data and discrimination in particular has highlighted how already socially unequal populations are discriminated against in data-driven systems. Elsewhere, there is growing interest in public perceptions of datafication, amongst academic researchers interested in producing 'bottom up' understandings of the new roles of data in society and non-academic stakeholders keen to establish positive perceptions of data-driven systems. However, research into public perceptions rarely engages with the issue of inequality which is so central in data and discrimination scholarship. Bringing these two issues together, this paper explores public perceptions of datafication through the lens of inequality, focusing on the relationship between understandings and feelings within these perceptions. The paper draws on empirical focus group research into how audiences perceive the data practices that signing in to access BBC digital services enable. The paper shows how inequalities relating to age, dis/ability, poverty and their intersections played a role in shaping perceptions and that these social inequalities informed understandings of and feelings about data practices in complex and diverse ways. It concludes with reflections on the significance of these findings for future research and for data-related policy.

ARTICLE HISTORY
Received 5 June 2019
Accepted 20 February 2020

KEYWORDS
Public perceptions; data practices; datafication; inequality; understanding; feelings

Introduction

The ubiquitous collection and use of digital data is changing people's lives, positively and negatively. Many commentators believe that it is vital that public views are factored into decisions that will shape the future of the data economy, and this has led to increasing interest in how the public perceives 'data practices' (that is, organisations collecting, analysing and sharing data and the outcomes of these processes). Research into public perceptions of datafication has therefore grown in recent years, and this has enhanced understanding of these matters. However, there are gaps in the research, which need to be addressed in order to advance understanding. This paper addresses those gaps.

CONTACT Helen Kennedy h.kennedy@sheffield.ac.uk

The first gap relates to inequality. A number of writers have highlighted how already socially unequal populations can be discriminated against in data-related practices. Examples include the high-profile *Propublica* revelations about how algorithmic systems used in criminal justice in the US discriminate on the grounds of race (Angwin et al., 2016) and a recent special issue of this journal on data justice. Yet research into people's *perceptions of* data practices – that is, what they understand and how they feel about them – rarely engages with this issue of inequality, something that is increasingly acknowledged in research into *experiences* of data practices – that is, how datafication affects lives.

The second gap relates to the relationship between the two elements of perceptions identified above: *understanding* (that is, knowledge and comprehension) and *feelings* (or emotional responses). Academic researchers have argued that feelings play an important role in shaping people's perceptions of data and related practices (Kennedy, 2018; Bucher, 2017). Policy debate is concerned with mechanisms to enhance public understanding, for example through data literacy initiatives (doteveryone, 2018b; Flyverbom, 2017), as it is assumed that better understanding will lead to more positive feelings about data practices. But the nature of the relationship between what people understand and their emotional responses to data practices has not been explored and remains unclear. Examining this relationship is important, to ascertain whether assumptions are well-founded and to advance knowledge.

Our paper contributes to debates about public perceptions of datafication by exploring the relationship between understanding and feelings about data practices in the context of social inequalities. It reports on empirical, focus group research which examined perceptions of the BBC's data practices amongst disadvantaged groups. The paper considers whether inequalities that have been identified as significant in relation to *experiences* of data practices are also important in relation to *understandings of* and *feelings about* them – these relate to race, class and gender (Eubanks, 2017; Noble, 2018). In our research, we found that age, dis/ability, poverty and their intersections played a role in shaping perceptions. Thus our paper highlights the importance of inequalities that have not been widely discussed to date in the literature on data and discrimination. It also highlights the complex and diverse ways in which social inequalities inform understandings of and feelings about data practices.

The media are an important site for exploring these matters. In the specific context of public service media (PSM), using data to deliver personalised services to audiences has become a key strategic priority across Europe (EBU, 2018). As a result, many PSM organisations are introducing the requirement that users sign in to a single digital account to enable such personalisation, including the BBC, which did so in 2018. As PSM organisations introduce datafied practices, their values (such universality and diversity) may inform what they do (Helberger, 2015), or they may be challenged by the commercial norms of the major global platforms that dominate online media provision. Questions of inequality concern the BBC, which seeks to include, reflect and represent the diversity of the UK in the services it offers. Thus in the paper we also consider whether PSM values influence diverse people's perceptions of the data practices of the BBC. In what follows, we discuss debates which inform our theoretical framework, outline our empirical research, and discuss our findings.

Perceptions of data practices and data-related inequalities

Researching perceptions of data practices

Recent years have witnessed a flurry of surveys and polls about people's understanding of data practices in the UK, where our study was based. For example, doteveryone found the majority of respondents to its survey knew that personal information is used to target advertising, but fewer realised their data could be sold to other companies or may determine the prices they are charged (2018a). A survey by Digital Catapult (2015) found that 96% of respondents claimed they understood the term 'personal data', but that less than two thirds chose the correct definition when presented with a series of options. Elsewhere, an Ipsos Global Trends survey (2016) found that 83% of UK respondents were unsure what information companies held about them.

As can be seen in the examples above, quantitative research focuses on what people know and understand about data practices rather than how they respond to them emotionally. Qualitative research, in contrast, has been more attentive to emotional responses to data practices. In research into how people feel about their health and self-tracking data, the work of Ruckenstein (2017) is one example. Bucher's (2017) small-scale study of how Facebook users imagine the platform's algorithms is another, this time in the context of social media platforms. The empirical research of one of us, Kennedy, has also highlighted the vital role that emotions play in how people make sense of data, arguing that data are 'as much felt as they are experienced cognitively and rationally' (2018, p. 831). Relatedly, one of D'Ignazio and Klein's (2016) six principles for inclusive, feminist-informed data visualisation is that it should 'legitimise affect,' or recognise affect as a legitimate way of knowing, comparable to comprehension-based ways of knowing.

These qualitative studies build on attempts to redeem the epistemological value of emotions from the historical and widely criticised reason/emotion binary (Bericat, 2016). There is an extensive literature which highlights the role emotions play in many aspects of life and which insists that emotions need to be taken seriously, as central aspects of social and cultural experience and as informing and informed by reason and rational thinking (Coleman, 2013). The entanglement of emotions with knowledge and reason means that they play a role in the formation of attitudes and the expression of perceptions, in relation to data practices as with other phenomena. We demonstrate this in our discussion of our empirical research below.

Existing qualitative and quantitative research has not explored this entanglement, or the relationship between understandings of and feelings about data practices more broadly. The doteveryone report concludes that 'without understanding it is likely that distrust of technologies may grow' (2018a,, p. 13), but this conclusion is not substantiated by the research on which the report is based, as the relationship between understanding and feelings of trust was not researched empirically. Some qualitative research draws conclusions about feelings without ascertaining whether participants actually understand the data practices that are the subject of the research, so not taking account of the understanding/feelings relationship. For example, Humphreys (2010) concludes that respondents in her study of a check-in app did not have strong feelings about institutional data-gathering through the app, because her respondents did not mention this issue. However, this may have resulted from a lack of knowledge and understanding rather than an

absence of strong feelings. Given assumptions in policy circles that better understanding of data practices will result in more positive emotional responses to them (doteveryone, 2018b; Flyverbom, 2017), it is vital that the relationship between understanding and feelings is subject to critical scrutiny.

A further limitation in research into public perceptions of data practices is that it rarely engages with the issue of inequality. One exception is a review by Understanding Patient Data (2018) of public attitudes to health data, which acknowledges demographic differences. The review notes that younger people are generally more knowledgeable about and supportive of data practices, that people in lower socio-economic groups are less likely to see the benefits of data practices and more likely to feel powerless to address data-related harms, and that ethnic minority groups are slightly less likely than ethnic majority groups to trust that their data will remain secure. This review aside, there is little attention to difference and inequalities in the research in this field. In the next section, we discuss literature that can contribute to a theoretical framework to help us to plug this gap.

Data and inequality

In recent years, researchers have nuanced generalised critiques of datafication's harms (such as increased surveillance, threats to privacy and new forms of algorithmic control) by pointing out that its negative effects are not equally felt by all (for example, Gangadharan & Niklas, 2019; Redden, 2018). Research into data and discrimination, much of which comes from the US, has highlighted how data practices disproportionately impact disadvantaged groups. Unlike the literature discussed in the previous section, which focuses on perceptions, data and discrimination scholarship focuses on experiences, or how the discriminatory consequences of widespread data mining affects lives.

One example is Eubanks' (2017) *Automating Inequality: how high-tech tools profile, police and punish the poor,* which highlights the impacts of data-driven discrimination on people living in poverty. Eubanks discusses examples of automated decision-making in which biases are written into data-driven systems. One way this happens in through proxies. For instance, in child abuse and neglect prediction, being hungry, not having enough clothes or living in a cold house – often conditions of poverty, as Eubanks points out – are taken as proxies for risk of child abuse or neglect.

Although primarily focused on poverty, Eubanks' research highlights intersections between class and race. Another proxy for whether a child is likely to be abused is if neighbours ring up to express concern. As statistics show that people are more likely to express concern about their black neighbours than their white neighbours, this proxy is interwoven with racial inequality. Other writers take up this focus on race and its intersections, such as Noble (2018), whose work focuses on the racial and gender bias of search engine algorithms. Noble presents a series of examples of how racist and sexist values are prioritised in decision-making tools thus 'masking and deepening inequalities' (2018, p. 1).

As these and other researchers persuasively argue, data practices discriminate. It is therefore important to take account of how social inequalities lead to unequal experiences of datafication. But do the same inequalities play a role in relation to *understandings of* and *feelings about* data practices? This paper addresses this important question, by considering a broader range of inequalities and intersections than data and discrimination literature has discussed to date. As evident in the examples above, much of this scholarship focuses

on race, perhaps reflecting its US origins, with some also focusing on the intersections of race with class and gender. Research on broader *digital* inequalities (for example, Yates et al., 2015) has identified how other inequalities are exacerbated in digital contexts, some of which were important in our research.

Age is one such factor. According to the UK's Office for National Statistics (ONS), of the 4.5 million adults who had never used the internet in 2018, more than half were aged 75 years and over (ONS, 2018). As societies become increasingly digital, this means that older people have unequal access to resources and services. Digitally excluded older people tend be less educated and more economically disadvantaged than those who are online (Robinson et al., 2015), a fact that highlights the intersection of age with class.

Disability is another important factor. According to Dobransky and Hargittai (2016), in the US, people with disabilities are less likely to use the internet than people without disabilities. When they do use it, they engage in a narrower range of activities than people who do not have disabilities. Age and class exacerbate this trend: people with disabilities who are older or with limited economic resources are less active internet users than younger and middle class people with disabilities. Poverty influences whether people with disabilities can access the internet, because of the cost of assistive technologies, for instance, meaning there are also important intersections between disability and class. Older people and people with disabilities have also been identified by the UK Government's Department for Culture, Media and Sport as significantly excluded from digital developments and therefore as priority groups in digital inclusion efforts such as the Digital Inclusion Fund launched in 2018 (Citizens Online, 2019).

Digital inequalities have their roots in broader inequalities, understood as 'the condition where people have unequal access to valued resources, services, and positions in society' (Kerbo, 2003, p. 11). In other words, digital inequalities are not only digital; rather, they are deep-seated and structural. Wyatt et al.'s (2002) study of internet non-users at the turn of the century, whilst not ostensibly focused on inequality, developed a typology of non-users which is useful in analysing digital and data inequalities, because it can help make sense of the relationship between how diverse populations use and understand digital and data-driven services and their feelings about them. It consists of: resisters (voluntary non-users), rejecters (former users), excluded (involuntary non-users), and expelled (involuntary former users). Our study highlighted that the boundary between resister and excluded is porous and that the two categories intersect, as we demonstrate below.

Inequalities also intersect, as intersectionality scholars have noted (Anthias & Yuval-Davis, 1983), and intersections also produce complex and diverse experiences. Disabilities are also extremely diverse, and people with distinct disabilities have very different experiences both of inequality and of engaging with digital and data-driven technologies. How we talk about inequality is similarly complex, and recent research into social class has proposed new, fine-grained classifications (Savage, 2015). In this paper, we bring this understanding of the complexity, diversity and intersectionality of social, digital and data inequalities together with a recognition that emotions, knowledge and reason work together as attitudes are formed and perceptions are expressed. This forms a framework with which we address the question: what do diverse audience groups understand and feel about signing in to access BBC services, the data this generates and the outcomes of this process? We elaborate on our findings and say more about the socially unequal groups to which our participants belong in the following sections.

Methods

The aim of our research was to explore understandings of and feelings about the BBC's data practices amongst people who are disadvantaged as a result of the social inequalities we discuss above. These are also groups that the BBC seeks to serve better, because of their unequal access to resources, services and positions in society (Kerbo, 2003). These include black, Asian, and minority ethnic (BAME) communities, people with disabilities, people with experiences of poverty, women, and both younger (defined by the BBC as 16–34) and older (over 65) people. To gather data about the demographic characteristics of our participants, we asked them to fill out a short questionnaire prior to the empirical research. We used this information along with participants' verbal self-descriptions in our analysis.

We used focus groups to carry out our research, for a number of reasons. First, focus groups allow access to diverse attitudes, feelings and beliefs. Second, when focus group members have a degree of homogeneity, as most of ours did, understanding of others' situations may facilitate discussion and group meaning-making (Krueger & Casey, 2009). To enable this, we recruited participants from pre-existing groups or communities where possible (see Table 1 for group composition). Third, when the topic discussed is one about which awareness may be limited, as in our study, participants can become informed about the topic through exposure to other group members' opinions and experiences. We carried out 11 focus groups with 68 participants in two cities in the north of England, chosen because of their comparatively low rates of sign in, according to internal BBC statistics, and significant levels of ethnic diversity and poverty.

We recruited participants for six of our focus groups by making contact with social, community and educational organisations for these groups. Two groups, the British Asian family and the older people's group, were recruited through one researcher's personal networks. Three groups were recruited through staff and student volunteer recruitment lists at two universities. Participants in seven groups knew each other prior to participating in the focus groups; those in the other four did not. We used personal networks and volunteer lists because funding constraints meant that we needed to complete the research in eight months, and we felt that these methods would ensure success in participant recruitment. Our resulting sample undoubtedly informs our findings – with more ethnic diversity amongst our participants, for example, our conclusions may be different. Table 2 summarises our participants' demographic characteristics.

Table 1. Focus group composition.

Location and composition	Numbers
Current and former University staff, control group	5
University staff, control group, all women	6
Creative technology centre for the unemployed	5
Arts centre for people with learning disabilities	5
Women's craft group, most older	12
Older people	6
British Asian family	4
Students at a 'new' university	2
Vocational training college	10
English as Second Language group	8
African Caribbean arts centre	7

Table 2. Participant demographic characteristics.

	Number of participants
Ethnicity	Asian British: 5 Asian: 2 Black African: 2 Black British: 3 Central and Eastern European: 3 Mixed/Multiple ethnic groups: 3 White British: 48 Other: 2
Disability	Mild Learning Disability (Attention Deficit Hyperactivity Disorder, Autism Spectrum Disorder, Asperger's): 10 Deaf: 2 (one also has MLD) Disability not specified: 14
Education level (only highest qualification listed)	No qualifications: 13 Secondary level qualifications: 34 Degree and professional qualification: 12 Higher degree and professional qualification: 7 Other/unknown: 2
Employment status	Employed, full time or part time, including self-employed: 18 Student: 17 Unemployed: 10 Retired: 19 Unknown/other (including carer): 4
Age	16–24: 19 25–34: 9 35–44: 9 45–54: 6 55–64: 6 65–74: 8 75+: 11
Gender	Women: 41, Non-binary: 1 Men: 26

Researching perspectives on data practices is difficult, because they are opaque by design and there is widespread confusion about them. We adopted a range of strategies to address this challenge. One of these was to use a cyclical approach. First, we identified what participants already knew about the BBC's data practices. Then, we provided them with information about these practices, adding to or clarifying what they already knew. We drew on publicly available information to do this. Finally, we explored how participants felt about the knowledge they acquired through participation in the focus groups.

We focused on data practices which result from signing in to iPlayer, the BBC's internet streaming and catch up service (for example, gathering demographic and media consumption data in order to personalise recommendations) and other common data practices (for example, social media analytics). We presented participants benefits that, according to the BBC, result from data gathering (such as personalisation), and concerns that were

expressed in online discussion groups when the requirement to sign in was introduced (for example, that data might be used to monitor TV licence payment). We tried to be balanced and neutral in what we said to participants, aware that methodological and research design choices can influence results (Law et al., 2011). Examples of what we said in the focus groups can be seen in Table 3, and a full transcript of our discussion guide can be found on our project website (https://livingwithdata.org/previous-research/signing-in-audience-experiences-of-data-practices-in-the-media/).

We adapted and explained questions in focus groups as required. For example, none of the participants in a focus group with people with learning disabilities used iPlayer, so we asked questions about data practices on platforms they did use, such as YouTube, to enable their participation in the discussion. Focus groups lasted between one and two hours. They were audio-recorded, transcribed and anonymised, after which audio-recordings were deleted. We gave participants a £20 one4all voucher to thank them for their contributions to our research. We secured ethical approval for our research from our university and complied with new European data protection regulation, the GDPR, when in came into force half way through our data-gathering phase. Data was organised and coded in Nvivo, according to pre-determined and emergent themes and categories, with particular attention paid to inequalities and their intersections.

Inequalities and perceptions of data practices

In this section, we discuss inequalities that appeared to be significant in informing our participants' understandings of and emotional responses to data practices: age, dis/ability, poverty and their intersections. While the majority of our participants used data-driven services such as iPlayer, YouTube or Spotify on a weekly basis, only around half of older participants (11/19) did so, compared to almost all young participants (27/28). Most participants who did not use data driven services weekly were older or had a disability, and one in three participants with experiences of poverty (4/12) did not use these services. We elaborate on the implications of these figures below.

Comparing younger and older participants' understanding and feelings about BBC data practices

As suggested above, data practices can be difficult to understand. This was the experience of most of the older participants in our research. Younger participants (16–34) had high levels of understanding, for example of why the BBC requires users to sign in and how it uses the data that signing in generates. However, there were also differences in

Table 3. Examples of what we told and asked participants about BBC data practices.

What's your understanding of why the BBC has started to require its users to sign in?
When you register for a BBC account you are asked to provide your gender and postcode. Why do you think the BBC wants this data?
The BBC collects data about you through your activity as you use their services (eg the programmes you watch, the channels you prefer, the types of media content you access, quizzes you do). Why do you think the BBC wants this data?
The BBC collects data about how you interact with their content on other platforms – eg, it monitors social media like Twitter or Facebook pages to see what its users are saying about it/its services. Why do you think the BBC wants this data?

understanding within this age group: young participants with mild learning disabilities struggled to understand the data practices that we discussed with them, as we highlight below.

Older participants (age 65 and above) displayed lower levels of understanding than younger participants. Around half did not use data-driven and personalised media. Of those that did use them, most used iPlayer exclusively (9/11). Some older people in our study felt left behind by data-driven developments, which they wanted but struggled to understand. Marilyn (75+, white British, has a disability, retired, iPlayer user) spoke passionately about her lack of understanding. She felt she was missing out on something that her children and grandchildren understood, stating 'you want to understand it. […] But you haven't got somehow the mental ability.' Knowledge and feelings intersected for Marilyn, who had a strong emotional response to her own lack of understanding.

Wyatt et al.'s (2002) typology can help make sense of our participants' perceptions of BBC data practices. For example, Marilyn could be described as 'excluded', or an involuntary non-user. Other older participants with limited understanding of BBC data practices were more like 'resisters', or voluntary non-users. For example, Jean (75+, white British, retired, iPlayer non-user) said 'as I don't do it [sign in to access digital services] it's nothing to do with me,' and Grace (65–74, black British, retired, iPlayer non-user) found personalised content annoying. Yet we found that the boundary between resistance and exclusion, or voluntary and involuntary non-use, is porous. This could be seen in Brenda's (65–74, white British, has a disability, retired, iPlayer non-user) response – she had not signed in because she 'wouldn't know what to do with it.' Her comments suggest that she was simultaneously excluded by a lack of skill, but also resistant because she did not see a purpose for personalised, data-driven services in her life. Resistance and exclusion alike were characterised simultaneously by knowledge, or lack thereof, and emotions, in all of these cases, we argue.

Most older participants felt that data-driven recommendations introduced unnecessary complications to media consumption, given that they had managed for many years without them. Larry (65–74, white British, unknown employment status, iPlayer user) felt that relying on recommendations was akin to living on microwave meals because 'you haven't got the imagination to cook.' In this instance, Larry was a 'resister' of this particular data-driven service. However, some older participants recognised the intersection of aging and disability, and this, combined with the BBC's public service remit to serve all groups in society, led them to believe that the BBC should offer data-driven personalisation, even if they would not use it themselves. Even Marilyn, the staunchest critic of personalisation in her group, said of older people with limited mobility 'if they know how to use it, they'd appreciate those services more than perhaps us more able bodied people do.' Participants in her group agreed that older people with limited mobility were often dependent on television for entertainment – George (75+, white British, retired, iPlayer user) said 'it's a lifeline for some people' – and that these people may be interested in using the broad range of digital services offered by the BBC, including personalisation. The public service value of universality and participants' awareness of the inequalities it is supposed to address influenced their reflections about BBC data practices.

Younger participants' greater knowledge of data-driven recommendation systems and the personalisation that they offer meant that they felt more positively about them than our older participants. The principal concern of most younger participants about signing

in was that they are required to do it too often. Few young participants expressed concern about the data practices that underlie the personal profiling on which recommendations are based. One exception was Tereza (16–24, white, Czech, employed, iPlayer non-user), who objected to personalised media because they collect unnecessary information. For Julia (25–34, white, Polish, employed, iPlayer former user), another exception, the public service remit of the BBC informed her views. She stopped using iPlayer once signing in became mandatory because she did not believe a public service broadcaster should collect data without users' informed consent. She did not feel the same way about commercial services and continued to subscribe to Netflix.

Young participants expressed little concern about the BBC collecting data about what they said about it on social media. Many younger participants felt that social media comments are public, and so expected the BBC to be 'on the ball,' as Jennifer (25–34, white British, employed, iPlayer user) put it, actively collecting these data. Michael (16–24, white British, university student, iPlayer user) saw posting about BBC services on Facebook or Twitter as a conscious act of opinion sharing and was more comfortable with the BBC mining data people shared in this way, than with data mining which is undertaken in contexts in which people 'passively forget about what [they're] providing.' Maya (16–24, mixed/multiple ethnic groups, British, unemployed, iPlayer user) was the only young participant who objected to social media data mining, because by aggregating data gathered from a range of platforms, the BBC would be 'building up a bigger picture of me than they need to.'

For young participants, their understandings of these varied data practices informed their feelings about them. In sharp contrast, the majority of older participants struggled to understand social media data mining and its purposes. Thus age played a significant role in informing participants' perceptions. In the absence of such understanding, older participants nonetheless expressed strong feelings. These emotional responses related to the broader data ecosystem in which particular data practices like social media data mining take place, and they were generally negative – for example, participants articulated feelings of suspicion, anxiety and fear. Shirley (75+, white British, retired, iPlayer user) opted out of using social media because she is 'afraid of it.' Peter (65–74, white British, retired, iPlayer user) said the BBC 'can manipulate' the data that it collects about its audiences, thus also expressing an ill-defined sense of suspicion about data mining practices. Marilyn was highly suspicious about what might happen to her data once it is in the data ecosystem, stating 'I don't trust anybody.' Deborah (75+, white British, retired, iPlayer user) felt the same. Responding to a question about whether she trusted the BBC to keep her data private, she declared 'you can't trust anybody about your information.' In these responses, limited understanding of data practices did not stop these older participants from having strong feelings which in turn influenced their attitudes, either to the specific data practices under discussion (as in Shirley's fear of social media data mining), what the BBC might do with user data (in Peter's case) or the datafied environment more broadly (Marilyn and Deborah). These comments challenge the assumption that understanding is a pre-requisite to forming views about data practices and show that emotions can inform perceptions in the absence of understanding. We saw a similar relationship between understanding, emotions and attitudes amongst young participants with mild learning disabilities, as we demonstrate below.

Nuancing the category of young people through the lens of disability

One of our focus groups took place at a vocational training college, where seven of the 10 participants had learning disabilities such as ADHD, ASD and Asperger's. In another focus group at an arts centre for people with learning disabilities, four of the five participants had similar disabilities. Learning disabilities often exist on a spectrum, on which our participants' disabilities were at the mild end – these are known as MLD (mild learning disabilities) (Emerson et al., 2001).

Young participants with and without MLD used data-driven, personalised media, except for one young participant with MLD, Ryan (16–24, white British, has MLD, unemployed, iPlayer non-user). Like some of our older participants, some young participants with MLD had reservations about data-driven services and the personalised content they make available. Brittany (25–34, white British, has MLD and deaf, unemployed, iPlayer non-user) explained that she did not like data-driven recommendations and preferred to look for her own videos than to follow recommendations. In commenting on his preferred entertainment platform, YouTube, Kyle (25–34, white British, has MLD, unemployed, iPlayer non-user) said that he generally turned off the videos that auto-play on YouTube after his selected videos have finished, because he likes to choose what he will watch himself. Using data-driven services selectively, these young participants were not excluded (that is, involuntary non-users) – rather, they were occasional resisters, voluntarily not using particular features of these services because of their feelings about them.

Participants with MLD did not understand some of the issues that we attempted to discuss with them in the focus groups, for example how personalised services work. Like our older participants, they also did not understand why media organisations mine social media data. They had limited understanding of specific data practices, but despite this, they had strong feelings about the outcomes of data mining. For example, at the time of research, upon registering to sign in to BBC services, users had to share their gender and their postcode. Our arts centre participants did not know why the BBC wanted these particular data, but they felt strongly about sharing them. Kyle, mentioned above, did not feel comfortable inputting his postcode, because he feared this would mean he could be geographically tracked. Ryan, also mentioned above, felt the same. He believed that the only data that the BBC needed from its audiences as they signed in was a password and their email address. Like some of the older participants in our study, younger people with MLD had largely negative feelings about the data ecosystem as a whole, and felt a need to protect themselves online. For example, Ryan was in favour of having to log on to platforms because 'it's better, better safe and no one knows your stuff.' Kyle said that if his privacy was breached and his personal data became public, 'I wouldn't be happy' and 'I'd be scared and angry to be honest.' Ryan and Brittany agreed.

Thus we can see that, although these young participants had the skills to engage with data-driven systems such as personalised media platforms, they had limited understanding of some data practices. But despite this, young participants with MLD had strong feelings, either about specific data practices (like gathering users' postcodes) or about the data ecosystem in which data practices take place. In their case, as with some of our older participants, understanding was not a pre-requisite for having strong feelings, which in turn influenced attitudes. The differences in understanding across young participants with and without MLD highlight the importance of taking an intersectional approach to the

relationship between social inequalities, understandings of and emotional responses to data practices. Doing so allows us to nuance the category of young people through the lens of disability.

How poverty informed perspectives on datafication

Income and education have been identified as factors which affect access to digital technologies in the UK. For example, the Oxford Internet Institute's Internet Survey of 2013 reports that limited access to digital technologies for people with low levels of education and experiences of poverty (as well as for older people) can exacerbate existing social inequalities, as more and more sectors of society become digital by default (OII, 2015). Similarly, according to the doteveryone survey (2018a) referenced above, wealthier people tend to see more benefit to digital and datafied technologies in their lives than people who are not wealthy – 57% compared to 43%. Exploring these phenomena in the context of datafication in the US, Eubanks (2017) persuasively argues that data-driven systems 'punish the poor.' In this context, it is not surprising that poverty appeared to be an important factor in our research. However, experiences of poverty appeared to impact participants' articulation of their perceptions of data practices in different ways to the other inequalities we have discussed thus far. We classify participants as living with poverty if they self-defined as unemployed, or talked about such experiences in our discussions with them.

It was striking how participants who were living with poverty navigated our conversations about BBC data practices. On a number of occasions, formerly vocal participants fell silent when we moved from talking about media use in general, of which they had experience, to talking about data-driven services, of which they did not. For example, Jason (35–44, white British, has a disability, unemployed, iPlayer non-user) did not have regular access to a personal computer or smart device, and therefore tended to access media in analogue form, for example by reading physical newspapers. He was at the creative technology centre for the unemployed where we carried out this focus group in order to use a computer, to which he would not otherwise have access. He contributed to the early stages of the focus group when the discussion was about media use, but as the discussion moved on to perceptions of data practices, he withdrew from the conversation.

Chris (55–64, mixed/multiple ethnic groups, British, non-binary, has MLD, unemployed, iPlayer non-user) arrived late to the same focus group as Jason, because he had had an appointment at the job centre. They described themselves as 'one of the poor mouses, you know, who's too poor to afford the TV license. So I don't really watch any television at all, unless at friends' houses.' They loved BBC radio and had passionate opinions about it, but not about data-driven services which they could not access in the absence of a TV license of personal computer. Chris responded to questions about media consumption and related data practices by talking instead about personal experience of poverty.

Older participants and younger participants with MLD articulated their perceptions of data practices from positions which often combined resister (voluntary non-user) and excluded (involuntary non-user) characteristics (Wyatt et al., 2002). A number of them suggested that they had made a choice not to engage. Even if this choice came from a position of limited understanding, they still felt that they were choosing. In contrast, Chris

and Jason were more clearly excluded, involuntary non-users because of their economic circumstances. They had not chosen not to engage. Their exclusion constrained their access to the understanding that comes from experience, and this seemed to have a silencing effect on these participants – they did not seem to feel comfortable expressing views about BBC data practices in the way that other participants with equally limited understanding did. Their silence seemed to come from a feeling of shame about their exclusion. This suggests that, for these participants, the structural inequalities that result from living with poverty impacted their confidence and willingness to develop and express views more strongly than being older or having MLD did.

As noted above, at the time of the research, audiences were requested to input their postcodes upon registering to sign in to BBC services. In online discussion fora, a common concern about this requirement was that these data could be used to enforce TV license payment in the future. We asked our participants whether they shared this concern, or whether they thought using sign in data to monitor license payment could be a good thing. Throughout the discussion at the older women's craft group, Brenda (65–74, white, female, has an unknown disability, retired, iPlayer non-user) made reference to her own limited economic resources. When the above question was asked, like Chris, Brenda responded to a question about data practices by talking about poverty, which, in her view, often led people to be unable to pay their TV licenses. She said: 'It's probably only all the young ones that are out of work [who aren't paying their TV licenses]. Where's the jobs for a start for them?' She felt that people do not pay their license fee because they cannot pay it. She did not care *how* TV licensing payment was monitored; rather, she cared about the relationship between its monitoring and inequality. With these comments Brenda suggested that her own experience of poverty and her awareness of how poverty impacts lives informed how she felt about data practices.

Virginia (75+, white British, retired, iPlayer user) expressed a similar view. She stated 'I don't object [to signing in] but then I can afford a [TV] license, so there's nobody going to come knocking on my door.' In Brenda and Virginia's comments, the public service remit of the BBC was important. These participants felt angry that limited access to devices and funds to pay a license fee excluded some people from engaging with BBC services, which are supposed to be for everyone. Feelings about BBC data practices were informed by consideration of what the BBC should do as a PSM, and these feelings intersected with participants' experiences of living with poverty, or their sympathy for people who do so. Moving the discussion about BBC data practices to a broader discussion about social inequality and poverty, these participants implicitly identified a relationship between data practices and social inequality, without necessarily articulating it as such.

Conclusions: advancing understanding of public perceptions of datafication through the lens of inequality

Our research makes a number of contributions to understandings of public perceptions of datafication. First, we found that social inequalities play a role in informing perceptions of data practices, in the same way that they have been shown to shape experiences of datafication. This finding is important for data policy as well as for data research. At the time of writing, new initiatives are being set up in the UK which aim to influence the governance of data uses, such as the government Centre for Data Ethics and Innovation (CDEI) and

the independent Ada Lovelace Institute (Ada), both of which claim that understanding public views will be at the heart of what they do. Understanding the role that social inequalities play in shaping such views is essential if both bodies are to achieve their aims – of ensuring that data work 'for people and society' (Ada's mission) and are 'a force for good' (a CDEI aim). Likewise, researchers interested in 'bottom up' understandings of datafication also need to take account of the structural inequalities that are entrenched by data-driven developments and how these relate both to how datafication is experienced and how it is perceived, in order to advance this field of research.

Second, our research explored whether the same inequalities that have been found to matter in relation to experiences of data-related discrimination were significant in relation to perceptions of data practices. Like Eubanks, we found that experiences of living with poverty were important. Research by Eubanks, Noble and others highlights how racism and sexism are written into data-driven systems and shape experiences of them. These factors did not surface in our study, but that is not to say that they do not matter in perceptions of datafication. Our findings are informed by both our sample and our focus – different participants and different questions about different data practices would produce different findings. Our research focused on comparatively mundane data practices; a focus on more contentious practices such as those discussed by Eubanks may also produce different results. We may have found, for example, that participants with experiences of living with poverty had experiences of and more to say about the datafied welfare state. These factors notwithstanding, our paper contributes to data studies scholarship by making visible the importance of inequalities that have not, to date, been discussed in the literature in this field. These inequalities relate to age and dis/ability. Being older and having a disability played a role in how participants perceived BBC data practices, from understandings of how data-driven recommendation systems work, to feelings about personalisation, and understandings of and emotional responses to data practices like social media data mining.

Thirdly, we found that social inequalities inform perceptions of data practices in complex and diverse ways. For some older participants, an absence of understanding did not necessarily mean they had no feelings about data practices. On the contrary, they had strong emotional responses to them, despite not understanding them. Younger participants with MLD understood data practices less well than their peers without MLD, but like some of our older participants, they also had strong feelings about them and their potential consequences. Without understanding the mechanisms behind data collection and sharing, they could still articulate clear feelings. These participants, who felt that they were resisting data practices rather than excluded from them, were comfortable developing and expressing their views.

However, if participants felt excluded, this was not the case. Falling silent when our focus group discussions turned to data-related issues, participants with experiences of poverty appeared to unwilling to speak about this topic. Feeling excluded from accessing data-driven services, rather than resisting them, seemed to have a silencing effect on these participants. Some participants with experiences of poverty responded to questions about data practices by commenting on the unequal economic conditions that lead people not to be able to make use of data-driven services, thus identifying a relationship between datafication and social inequality. Poverty demarcated and was exclusionary, more so than other inequalities that surfaced in our study. Changing the conversation about perceptions of data practices to 'connect data back to the social and political reality from

which they were produced' (D'Ignazio & Bhargava, 2019, p. 131) as some of our participants did could be seen as decentring of data, which Gangadharan and Niklas (2019) propose is necessary in order to understand the role of datafication in the production of inequalities. In other words, by prioritising economic and social conditions, these participants acknowledged datafication's connection to larger systems of structural inequality. We say more about decentring below.

Fourthly, our research shows how understandings of public service values and the inequalities they are intended to address related to perceptions of BBC data practices amongst some of our participants. A number of participants expressed the view that the public service remit of the BBC to serve all groups in society should inform its data practices. For some, this meant it should offer data-driven personalisation, as this could act as a 'lifeline' for people with mobility difficulties. Others 'decentred' BBC data practices in our discussions, expressing anger that BBC services, which were supposed to be for everyone, are actually inaccessible to people without digital devices or the funds to pay a license fee. These participants' comments show how the specific contexts of datafication – in this case, a PSM context – play a role in public perceptions of data practices. This highlights the need to move beyond generalised claims about datafication and instead ground analyses in specific settings in order to 'develop understanding of the material contexts in which datafication has effects' (Kennedy & Hill, 2017, p. 702).

Finally, our research challenges the assumption that understanding is the main prerequisite to developing views about data practices. Many of our older participants and participants with MLD based their views of data practices on their feelings rather than their understanding. This demonstrates the role emotions play in shaping perceptions of datafication and the need to take emotions seriously, as informing the formation of attitudes and the expression of perceptions, in relation to data practices as with other phenomena.

Older participants and participants with MLD had strong, often negative, emotional reactions. One possible conclusion from this finding is that to reduce negative feelings, we need more understanding. However, in our study we found that not all participants *can* comprehend the complexities of datafication – how personalisation works and the purpose of social media data mining are examples of data practices that a number of participant groups struggled to understand. In this context, data literacy initiatives which aim only to increase understanding will not necessarily have positive results. Our research suggests that feelings offer a way to engage with some people about datafication. This point has also been made by other scholars concerned with the relationship between structural inequalities and datafication. D'Ignazio and Bhargava (2019) have found that emotional connections facilitate engagement in data projects they have organised – it is for this reason that one of D'Ignazio and Klein's (2016) six principles for inclusive data visualisation is that it should 'legitimise affect'. These authors argue that Patricia Hill Collins's (2009, p. 266) proposition that 'neither emotion nor ethics is subordinated to reason' in ideal forms of knowledge holds true in this context. Working with emotional responses in projects which seek to engage publics on data-related issues, as D'Ignazio and Bhargava do, might lead us towards a more inclusive datafied society, at least for some groups, in some contexts.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Arts and Humanities Research Council [grant number AH/R013322/1].

Notes on contributors

Helen Kennedy is professor of digital society at the University of Sheffield and director of the Living With Data programme of research hosted there (https://livingwithdata.org/).

Robin Steedman is a postdoctoral researcher interested in diversity and inequality in the media, formerly of the University of Sheffield and now of Copenhagen Business School.

Rhianne Jones is a research lead at BBC Research & Development where her research focuses on examining the development and use of data-driven technology and the public interest.

Bibliography

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. *ProPublica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Anthias, F., & Yuval-Davis, N. (1983). Contextualising feminism: Gender, ethnic and class divisions. *Feminist Review*, *15*(1), 62–75. https://doi.org/10.1057/fr.1983.33

Bericat, E. (2016). The sociology of emotions: four decades of progress. *Current Sociology*, *64*(3), 491–513. https://doi.org/10.1177/0011392115588355

Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society*, *20*(1), 30–44. https://doi.org/10.1080/1369118X.2016.1154086

Citizens Online. (2019, June 3). Objectives. https://www.citizensonline.org.uk/digital-inclusion-fund/

Coleman, S. (2013). *How voters feel*. Cambridge University Press.

Collins, P. H. (2009). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment* (2nd ed. Routledge.

Digital Catapult. (2015). *Trust in personal data: A UK review*. Report by Digital Catapult.

D'Ignazio, C., & Bhargava, R. (2019). Data visualization literacy: A feminist starting point. In M. Engebretsen & H. Kennedy (Eds.), *Data Visualization in society* (pp. 207–222). Amsterdam University Press.

D'Ignazio, C., & Klein, L. (2016, October 23–28). *Feminist data visualization* [Paper presentation]. Workshop on Visualization for the Digital Humanities, IEEE VIS Conference, Baltimore, Maryland, United States.

Dobransky, K., & Hargittai, E. (2016). Unrealized potential: Exploring the digital disability divide. *Poetics*, *58*, 18–28. https://doi.org/10.1016/j.poetic.2016.08.003

doteveryone. (2018a). *People, power and technology: The 2018 digital attitudes report*. https://understanding.doteveryone.org.uk

doteveryone. (2018b). *People, power, and technology: The 2018 digital understanding report*. https://understanding.doteveryone.org.uk

Emerson, E., Hatton, C., Felce, D., & Murphy, G. (2001). *Learning disabilities – the fundamental facts*. The Foundation for People with Learning Disabilities.

Eubanks, V. (2017). *Automating inequality: How high-tech tools profile, police and punish the poor*. St Martins Press.

European Broadcasting Union (EBU). (2018). *Big data initiative: Activity report 2017–18*. https://www.ebu.ch/publications/big-datat-initiative—activity-report-2017-2018

Flyverbom, M. (2017). Datafication, transparency and trust in the digital domain. In *Trust at risk: Implications for EU policies and institutions* (pp. 69–84). European Commission Report. https://openarchive.cbs.dk/bitstream/handle/10398/9539/Final%20published%20report.pdf?sequence=1

Gangadharan, S. P., & Niklas, J. (2019). Decentering technology in discourse on discrimination. *Information, Communication and Society*, *22*(7), 882–899. https://doi.org/10.1080/1369118X.2019.1593484

Helberger, N. (2015). Merely facilitating or actively stimulating diverse media choices? Public service media at the crossroad. *International Journal of Communication*, *9*, 1324–1340. https://hdl.handle.net/11245/1.519098

Humphreys, L. (2010). Mobile social networks and urban public space. *New Media & Society*, *12*(5), 763–778. https://doi.org/10.1177/1461444809349578

Ipsos Global Trends. (2016). Awareness of personal information held by companies. https://www.ipsosglobaltrends.com/awareness-of-personal-information-held-by-companies/

Kennedy, H., & Hill, R. (2017). 'The feeling of numbers: emotions in everyday engagements with data and their visualisation', *Sociology*. http://journals.sagepub.com/doi/full/10.1177/0038038516674675

Kennedy, H. (2018). 'Living with data: Aligning data studies and data activism through a focus on everyday experiences of datafiction', *Krisis: journal for contemporary philosophy*. https://krisis.eu/living-with-data/

Kerbo, H. R. (2003). *Social stratification and inequality: Class conflict in historical and global perspective*. McGraw-Hill.

Krueger, R., & Casey, M. A. (2009). *Focus groups: A practical guide for applied research*. Sage.

Law, J., Ruppert, E., & Savage, M. (2011). The double social life of methods. *CRESC Working Paper*, 95. Retrieved June 4, 2012, from http://www.open.ac.uk/researchprojects/iccm/library/164

Noble, S. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

Office for National Statistics (ONS). (2018). Statistical bulletin: Internet users, UK: 2018, May. https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2018

Open Data Institute (ODI). (2018). Who do we trust with personal data? July. https://theodi.org/article/who-do-we-trust-with-personal-data-odi-commissioned-survey-reveals-most-and-least-trusted-sectors-across-europe/

Oxford Internet Institute (OII). (2015). The Oxford Internet surveys: Internet use, behaviour and attitudes in Great Britain 2003–2015. http://blogs.oii.ox.ac.uk/oxis/wp-content/uploads/sites/43/2015/01/OxIS-Brochure.pdf

Redden, J. (2018). The harm that data do. *Scientific American*, *319*(5). https://www.scientificamerican.com/article/the-harm-that-data-do/

Robinson, L., Cotton, S. R., Ono, H., Quan-Haase, A., Mesch, G., Chen, W., Schulz, J., Hale, T. M., & Stern, M. J. (2015). Digital inequalities and why they matter. *Information, Communication & Society*, *18*(5), 569–582. https://doi.org/10.1080/1369118X.2015.1012532

Ruckenstein, M. (2017). Keeping data alive: talking DTC genetic testing. *Information, Communication & Society*, *20*(7), 1024–1039. https://doi.org/10.1080/1369118X.2016.1203975

Savage, M. (2015). *Social class in the 21st century*. Pelican.

Understanding Patient Data. (2018). Public attitudes to patient data use: A summary of existing research. http://understandingpatientdata.org.uk/sites/default/files/2019-01/Public%20attitudes%20key%20themes_0.pdf

Wyatt, S., Thomas, G., & Terranova, T. (2002). They came, they surfed, the went back to the beach: Conceptualizing use and non-use of the internet. In S. Woolgar (Ed.), *Virtual society? Technology, cyberbole, reality* (pp. 23–40). Oxford University Press.

Yates, S., Kirby, J., & Lockley, E. (2015). Digital media use: differences and inequalities in relation to class and age. *Sociological Research Online*, *20*(4), 71–91. https://doi.org/10.5153/sro.3751